

УДК 004.021

## АЛГОРИТМ ДЛЯ КРОСС-ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ

Дмитриева И.А. (Национальный исследовательский университет ИТМО)  
Научный руководитель – к.т.н., научный сотрудник ФБИТ Виксинн И.И.  
(Национальный исследовательский университет ИТМО)

В работе были проанализированы существующие методы кросс-идентификации пользователей, позволяющие идентифицировать пользователей в двух социальных сетях, разработан модифицированный алгоритм, повышающий точность кросс-идентификации. Были определены наиболее заполняемые параметры в профилях, по которым следует проводить кросс-идентификацию и повышена точность алгоритма кросс-идентификации пользователей социальных сетей.

**Введение.** В современном мире социальные сети занимают достаточно большую часть жизни людей. Благодаря социальным сетям люди могут общаться, не выходя из дома, выкладывать фотографии, обмениваться мультимедийным контентом. Существует множество социальных сетей, каждая из которых имеет свои особенности: Twitter – позволяет обмениваться статусами и короткими сообщениями, LinkedIn – создан для обмена рабочими контактами и для профессиональной деятельности, Instagram – позволяет обмениваться сообщениями, выкладывать видео и фотографиями, а Вконтакте и Facebook – предоставляют возможность создания пользовательских групп и сообществ по интересам, планирования встреч и мероприятий, проведения прямых трансляций, обмена сообщениями, звонками и блогами. В данной работе мы как раз фокусируемся на идентификации пользователей на основе их дружеских связей и атрибутах данных профилей, и для этого используем Facebook и Вконтакте.

**Основная часть.** В ходе исследования существующих алгоритмов кросс-идентификации было выявлено, что наиболее популярным методом кросс-идентификации является сравнение на основе атрибутов профилей социальных сетей. Однако этот метод ограничен тем, что некоторые социальные сети слишком различны между собой. Поэтому в целях улучшения качества оценки к атрибутам профиля добавляют различные дополнительные параметры сравнения, такие как: социальные отношения, контент пользовательских записей, анализ описания профиля, распознавание лиц на фотографиях и проверку ссылок, представленных в контенте профиля.

Метод кросс-идентификации пользователей, использующийся в данной статье основан на методе, базируется на комбинации взаимоотношений пользователей – друзьях и параметров профилей пользователей в социальных сетях. Данный метод основывается на том факте, что пользователи предпочитают поддерживать дружественные связи в различных социальных сетях, собирая в них своих родственников, коллег, школьных друзей и сокурсников. Переходя из одной социальной сети в другую, пользователи частично переносят свои дружеские сети из первой во вторую. Таким образом формируется некоторый круг друзей, присутствующих в разных сетях, по которым появляется возможность идентифицировать начального пользователя.

В ходе разработки алгоритма было проведено дополнительное исследование социальной сети Facebook, которое показало какие параметры профиля пользователи заполняют больше всего, а какие практически не заполняют. На основе этого исследования были выбраны параметры пользователя, по которым будет проводиться кросс-идентификация при помощи разработанного модифицированного алгоритма.

**Выводы.** В ходе работы были исследованы существующие решения по кросс-идентификации пользователей социальных сетей и исследована социальная сеть Facebook. Также был

разработан модифицированный алгоритм кросс-идентификации пользователей на основе атрибутов профиля и социальных связей пользователя.

Дмитриева И.А. (автор)

Подпись

Виксин И.И. (научный руководитель)

Подпись